PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-356799

(43)Date of publication of application : 26.12.2001

(51)Int.Cl. G10L 21/04
G10K 15/04
G10L 13/00
H03M 7/30

(21)Application number : 2000-175065

(22)Date of filing : 12.06.2000

(71)Applicant : TOSHIBA CORP

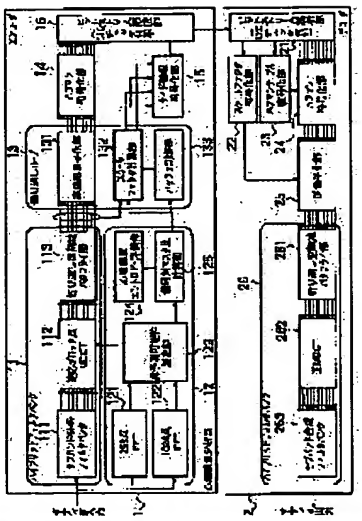
(72)Inventor : OKAZAKI MASAHIKO
KOJIMA YOSHINARI
WAKASUGI JUN

(54) DEVICE AND METHOD FOR TIME/PITCH CONVERSION

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a device and a method for time/pitch conversion which can easily vary the pitch and reproduction time of a reproduced voice without bringing large-sized constitution nor complexity of processing nor spoiling reproduced sound quality.

SOLUTION: After the spectrum of voice data compressed as frequency data is shifted, the data are interpolated and thinned out and reconverted into voice data of time-series data.



LEGAL STATUS

[Date of request for examination] 28.01.2003

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration] abandonment

[Date of final disposal for application] 06.01.2004

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

CLAIMS

[Claim(s)]
[Claim 1] They are the time / pitch inverter provided in the voice regeneration system which inputs the voice data compressed as frequency data, carries out inverse transformation of the voice data compressed as frequency data from a frequency domain to a time domain, and obtains the voice data of time series data. In case inverse transformation of the voice data compressed as frequency data is carried out from a frequency domain to a time domain and the voice data of time series data is obtained A shift means to shift the spectrum of the voice data in a frequency domain according to the pitch converted quantity of voice data, and to determine the playback frequency of the voice data of time series data, Voice data is interpolated or operated on a curtailed schedule to the spectrum in the frequency domain which was shifted by said shift means and obtained. It has interpolation/infanticide means which makes the same the number of voice data of the spectrum in the frequency domain before and behind a shift with the same bandwidth. The time / pitch inverter characterized by changing the pitch of voice data in case inverse transformation of the voice data in the frequency domain obtained with said interpolation/infanticide means is carried out to the voice data of time series data.
[Claim 2] The voice data compressed as frequency data is inputted. They are the time / pitch inverter provided in the voice regeneration system which changes into analog voice data the digitized voice data of the time series data obtained in the voice data compressed as frequency data by carrying out inverse transformation from a frequency domain to a time domain by DAC, and is reproduced. In case inverse transformation of the voice data compressed as frequency data is carried out from a frequency domain to a time domain and the voice data of time series data is obtained A shift means to shift the spectrum of the voice data in a frequency domain according to the playback time amount of playback voice, and to determine the playback frequency of the voice data of time series data, Interpolation/infanticide means which interpolates or operates voice data on a curtailed schedule to the spectrum in the frequency domain which was shifted by said shift means and obtained, and makes the same the number of voice data of the spectrum in the frequency domain before and behind a shift with the same bandwidth, According to the playback time amount of playback voice, a frequency generates an adjustable clock signal. It has a clock generation means to supply the generated clock signal to said DAC at least. The time / pitch inverter characterized by extending / shortening the playback time amount of voice data in case said DAC changes the digitized voice data of time series data into analog voice data based on the clock signal supplied from said clock generation means.
[Claim 3] The voice data compressed as said frequency data is the time / pitch inverter according to claim 1 or 2 characterized by being stored in the storage media in which the data read-out rate of arbitration is possible.
[Claim 4] In case the voice data compressed as frequency data is inputted, inverse transformation of the voice data compressed as frequency data is carried out from a frequency domain to a time domain and the voice data of time series data is obtained According to the pitch variation of voice data, the spectrum of the voice data in a frequency domain is shifted. Voice data is interpolated or operated on a curtailed schedule to the spectrum in the frequency domain which determined the playback frequency of the voice data of time series data, was shifted and was obtained. The time / the pitch conversion approach characterized by changing the pitch of voice data in case the number of voice data of the spectrum in the frequency domain before and behind a shift is made the same with the same bandwidth and inverse transformation of the voice data in the frequency domain obtained by interpolation/infanticide is carried out to the voice data of time series data.

[Claim 5] In case the voice data compressed as frequency data is inputted, inverse transformation of the voice data compressed as frequency data is carried out from a frequency domain to a time domain and the voice data of time series data is obtained According to the playback time amount of playback voice, the spectrum of the voice data in a frequency domain is shifted. Voice data is interpolated or operated on a curtailed schedule to the spectrum in the frequency domain which determined the playback frequency of the voice data of time series data, was shifted and was obtained. The number of voice data of the spectrum in the frequency domain before and behind a shift is made the same with the same bandwidth. According to the playback time amount of playback voice, a frequency generates an adjustable clock signal. The digitized voice data of the time series data which supplied the generated clock signal to DAC at least, and were obtained from the frequency domain by the inverse transformation to a time domain The time / the pitch conversion approach characterized by extending / shortening the playback time amount of voice data in case it changes into analog voice data based on the clock signal with which said DAC was supplied.

[Translation done.]

DETAILED DESCRIPTION

[Detailed Description of the Invention]
[0001]
[Field of the Invention] This invention relates to the time / pitch inverter, and the time / the pitch conversion approach of performing the time of playback voice, or pitch conversion in the system which reproduces the signal whose input is not time series data but frequency data.
[0002]
[Description of the Prior Art] The technique of pitch conversion is needed for various applications, such as pitch controllers, such as speech rate inverters, such as equipment which changes performance time amount, such as an effector for pitch conversion of recording, and commercial work, a minutes sound, an interview, and news, and karaoke.
[0003] Conventionally, it is divided roughly into two kinds, processing in a time domain, and processing in a frequency domain, as the technique of changing the pitch of voice data. In processing in a time domain, on the time-axis, the wave-like break point occurred and it had appeared as a jarring noise at the time of voice playback. Since there was no generating of such a break point at processing in a frequency domain compared with this, a noise was not generated. However, by media, such as a tape and CD, since voice is recorded as time series data, in order to perform pitch conversion in a frequency domain, time amount <—> frequency conversion, such as FFT (fast Fourier transform), needed to be performed. However, many had to be calculated to perform FFT and it had the fault that the throughput of an arithmetic circuit had to be large.
[0004] Next, pitch conversion is explained to a detail.
[0005] The latter technique was used for the system with a demand mainly severe in simple systems, such as key control of karaoke, concerning [the former technique] the tone quality of a musical instrument etc., although

were based on data processing in the thing (b) frequency domain depended on data processing in the (a) time domain and it was divided roughly into two kinds as the technique of changing a pitch, as mentioned above.

[0006] An example of the pitch conversion by the technique of the above (a) is shown in drawing 13 . In processing in a time domain, although a rise/down of a pitch are performed by controlling the reproduction speed of time series data, as shown in drawing 13 , cautions are required for playback time amount to be shortened or extended by coincidence. That is, when a pitch is lowered, playback time amount is extended by coincidence, and when a pitch is raised on the other hand, playback time amount is shortened by coincidence. Here, playback time amount is not changed, but it aims at changing only a pitch, and playback time amount must be the same as it of former data. Therefore, when a duplication part surely arises in somewhere when the pitch of former data is lowered, and a pitch is raised, the lack part of data will surely arise in somewhere. Since these become discontinuous [the data on time series], if it reproduces as it is, a noise will occur and tone quality will worsen. There is cross fade processing as a technique for avoiding such fault. This processing carries out fade-out of the termination of a continuous wave form, when a pitch is lowered, as shown in drawing 14 , it carries out fade-in of the initiation of the following continuous wave form to it and coincidence, and performs cross fade continuation. The noise in a node decreases by this. On the other hand, when a pitch is raised, in order to compensate the lack part of data, the same data are reproduced twice, and the noise in a node decreases by cross fade continuation similarly. However, in this cross fade processing, when the phase of a fade-out sound and a fade-in sound is reversed, a good result may be unable to be obtained. Moreover, it was also regarded as questionable that a periodic wave occurs in a playback sound.

[0007] Next, the technique of changing a pitch by processing of the above (b) can perform pitch change easily in shifting data on a frequency shaft, as shown in drawing 15 , and the break point on a time-axis does not generate it, either. For this reason, compared with the above (a), the tone quality of the description of a playback sound is good. However, the voice data outputted from a tape, CD, etc. is time series data, and in order to change this into a frequency domain from a time domain, it needs data processing, such as FFT. Although the equipment or the systems which mainly consist of an arithmetic circuit and memory, such as DSP (digital signal processor), could perform this data processing, many had to be calculated and there was a fault that the throughput of an arithmetic circuit had to be large.

[0008] Next, the time conversion technique of changing the playback time amount of voice data is explained.

[0009] It is used for the device which calls it a time stretch / compression to perform only compaction of playback time amount, and extension, without changing the pitch of a playback sound, and is mainly called speech rate conversion and sampler. This applies the technique of the pitch conversion mentioned above, and is realizable.

[0010] Since it mentioned above and the pitch of a playback sound falls when reproduction speed is made late and playback time amount is lengthened, it is operated so that this may be returned to the original pitch using the technique of pitch conversion. Thereby, as shown in drawing 16 , a pitch remains as it is and can extend only playback time amount. What is necessary is just to perform actuation contrary to this for on the other hand shortening playback time amount.

[0011] when the media which recorded time series data use well until now , such as CD and a music tape , as they were be reproduced and a time stretch / compression be performed , the read-out rate from media be made adjustable using the equipment which control reproduction speed, or reproduction speed remained as it was and the technique of give big buffer memory to a system and adjust playback time amount be adopted . However, it did not result until the complicated additional equipment and large-scale processing were needed and both could be realized easily.

[0012]

[Problem(s) to be Solved by the Invention] As explained above, the fault that it was difficult to remove a noise from a playback sound certainly even if it performs this processing, and tone quality deteriorated although cross fade processing for avoiding the discontinuity of voice data in processing in a time domain is performed was caused among the conventional conversion technique of changing the pitch of voice data. In order the processing

which changes voice data into a frequency domain from a time domain is needed in processing in a frequency domain on the other hand and to perform this processing, the fault that a large-scale configuration and great time amount were needed was caused.

[0013] Then, this invention is made in view of the above, and the place made into the purpose is to offer the time / pitch inverter, and the time / the pitch conversion approach of changing easily the pitch / playback time amount of playback voice, without [without it causes enlargement of a configuration, and complication of processing, and] spoiling playback tone quality.

[0014]

[Means for Solving the Problem] In order to attain the above-mentioned purpose, 1st means to solve a technical problem They are the time / pitch inverter provided in the voice regeneration system which inputs the voice data compressed as frequency data, carries out inverse transformation of the voice data compressed as frequency data from a frequency domain to a time domain, and obtains the voice data of time series data. In case inverse transformation of the voice data compressed as frequency data is carried out from a frequency domain to a time domain and the voice data of time series data is obtained A shift means to shift the spectrum of the voice data in a frequency domain according to the pitch converted quantity of voice data, and to determine the playback frequency of the voice data of time series data, Voice data is interpolated or operated on a curtailed schedule to the spectrum in the frequency domain which was shifted by said shift means and obtained. It has interpolation/infanticide means which makes the same the number of voice data of the spectrum in the frequency domain before and behind a shift with the same bandwidth. In case inverse transformation of the voice data in the frequency domain obtained with said interpolation/infanticide means is carried out to the voice data of time series data, it is characterized by changing the pitch of voice data.

[0015] The 2nd means inputs the voice data compressed as frequency data. They are the time / pitch inverter provided in the voice regeneration system which changes into analog voice data the digitized voice data of the time series data obtained in the voice data compressed as frequency data by carrying out inverse transformation from a frequency domain to a time domain by DAC, and is reproduced. In case inverse transformation of the voice data compressed as frequency data is carried out from a frequency domain to a time domain and the voice data of time series data is obtained A shift means to shift the spectrum of the voice data in a frequency domain according to the playback time amount of playback voice, and to determine the playback frequency of the voice data of time series data, Interpolation/infanticide means which interpolates or operates voice data on a curtailed schedule to the spectrum in the frequency domain which was shifted by said shift means and obtained, and makes the same the number of voice data of the spectrum in the frequency domain before and behind a shift with the same bandwidth, According to the playback time amount of playback voice, a frequency generates an adjustable clock signal. It has a clock generation means to supply the generated clock signal to said DAC at least. In case said DAC changes the digitized voice data of time series data into analog voice data based on the clock signal supplied from said clock generation means, it is characterized by extending / shortening the playback time amount of voice data.

[0016] Voice data with which the 3rd means was compressed as said frequency data in said 1st or 2nd means is characterized by being stored in the storage media in which the data read-out rate of arbitration is possible.

[0017] The 4th means inputs the voice data compressed as frequency data. In case inverse transformation of the voice data compressed as frequency data is carried out from a frequency domain to a time domain and the voice data of time series data is obtained According to the pitch variation of voice data, the spectrum of the voice data in a frequency domain is shifted. Voice data is interpolated or operated on a curtailed schedule to the spectrum in the frequency domain which determined the playback frequency of the voice data of time series data, was shifted and was obtained. In case the number of voice data of the spectrum in the frequency domain before and behind a shift is made the same with the same bandwidth and inverse transformation of the voice data in the frequency domain obtained by interpolation/infanticide is carried out to the voice data of time series data, it is characterized by changing the pitch of voice data.

[0018] The 5th means inputs the voice data compressed as frequency data. In case inverse transformation of the

voice data compressed as frequency data is carried out from a frequency domain to a time domain and the voice data of time series data is obtained According to the playback time amount of playback voice, the spectrum of the voice data in a frequency domain is shifted. Voice data is interpolated or operated on a curtailed schedule to the spectrum in the frequency domain which determined the playback frequency of the voice data of time series data, was shifted and was obtained. The number of voice data of the spectrum in the frequency domain before and behind a shift is made the same with the same bandwidth. According to the playback time amount of playback voice, a frequency generates an adjustable clock signal. The digitized voice data of the time series data which supplied the generated clock signal to DAC at least, and were obtained from the frequency domain by the inverse transformation to a time domain In case it changes into analog voice data based on the clock signal with which said DAC was supplied, it is characterized by extending / shortening the playback time amount of voice data.

[0019]

[Embodiment of the Invention] Hereafter, 1 operation gestalt of this invention is explained using a drawing.

[0020] Drawing 1 shows the configuration of an MP3 encoder / decoder including the function of the time / pitch inverter concerning 1 operation gestalt of this invention.

[0021] This operation gestalt explains the pitch conversion at the time of reproducing the compression voice compressed by the MP3 method which is one of the MPEG speech compression methods. In addition, since all are applicable if voice data is frequency data, besides MP3, even if it is MPEG speech compression methods, such as ACC, it can carry out, and especially speech compression is not limited to an MPEG method. Since the compression voice data based on MPEG is already recorded as frequency data, it does not have to carry out a frequency and time amount conversion like playback of the media which recorded time series data. This point is used, spectrum information on a frequency domain is operated only by adding the program of a number step to the software which performs the algorithm of filter data processing, without changing most filter data processing further performed at the time of decoding of the compression voice data of MPEG, and it is made to realize pitch conversion of playback voice easily.

[0022] In drawing 1 , the MP3 encoder / decoder of this operation gestalt input the voice data which is time series data, is equipped with the encoder 1 which carries out compression conversion of this voice data with the compression method of MP3 known from the former at the data in a frequency domain, and the decoder 2 which carries out inverse transformation of this output to time series data, and is outputted as voice data of time series data in response to the output in the frequency domain of this encoder 1, and is constituted. An encoder 1 The hybrid filter bank 11 and the mental acoustic-sense analyzor 12, The repeat loop formation 13 and the Huffman coding section 14 which performs Huffman coding processing in response to the output of the repeat loop formation 13, The side information coding section 15 which encodes side information in response to the output of the repeat loop formation 13, It has the bit stream formation section 16 which forms a bit stream in response to the output of the Huffman coding section 14, and the output of the side information coding section 15. The hybrid filter bank 11 It has the clinch distorted reduction butterfly section 113 with the subband analysis filter bank 111 and the adaptation block length MDCT112. The mental acoustic-sense analyzor 12 FET (fast Fourier transform)121 of 256 points, and FFT122 of 1024 points, It has the predictability-ed test section 123, the mental acoustic-sense entropy evaluation section 124, and the signal pair mask ratio count section 125, and the repeat loop formation 13 is equipped with the nonlinear quantization section 131, the scale-factor count section 132, and the buffer control section 133, and is constituted.

[0023] The bit stream analysis section 21 in which a decoder 2 analyzes a bit stream in response to the output in the frequency domain of the bit stream formation section 16 of an encoder 1, The scale-factor decryption section 22 which performs a scale-factor decryption in response to the output of the bit stream analysis section 21, The Huffman table decryption section 23 which performs the Huffman table decryption in response to the output of the bit stream analysis section 21, The Huffman coding section 24 which performs Huffman coding in response to the output of the bit stream analysis section 21 and the Huffman table decryption section 23, The reverse quantization section 25 which performs reverse quantization and acquires spectrum information in response to the output of the scale-factor decryption section 22 and the Huffman coding section 24, In response to the

output of the reverse quantization section 25, the voice data as time series data is reproduced. And it has the hybrid filter bank 26 including the shift means and interpolation/infanticide means of performing pitch transform processing which serves as the description of this operation gestalt in this renewal process. The clinch distorted reduction butterfly section 261 which carries out the butterfly session of the spectrum information from which the hybrid filter bank 26 was obtained in the reverse quantization section 25, In response to the output of the clinch distorted reduction butterfly section 261, in response to the output of reverse MDCT262 which performs an inverse Fourier transform, and reverse MDCT262, it has the subband composition filter bank 263 which performs subband composition, and is constituted.

[0024] In the hybrid filter bank 26 of a decoder 2, although processing of butterfly session, reverse MDCT, and QMF composition is performed, these processings are processed as one collected algorithm by software. Moreover, with this algorithm, in order to perform pitch transform processing, in case a shift means performs a frequency and time amount conversion first, spectrum information on a frequency domain is shifted, the frequency of playback voice is determined, interpolation of the data in a frequency domain or processing of infanticide is performed to the spectrum information shifted with interpolation/infanticide means, and the number of data is arranged. While changing a pitch, when spectrum information is returned to a time domain by this, it is made for playback time amount not to change.

[0025] Next, they are explained with reference to drawing 3 – drawing 9 about the above-mentioned processing, using the sinusoidal data of a frequency domain as shown in drawing 2 as an example. Hereafter, it explains based on the result which carried out simulation about the spectrum information on 0–16kHz of bands using FFT / reverse FFT. The data inputted into reverse FFT are set to a 1kHz sine wave, sampling frequency =32kHz, and measurement size =64.

[0026] In not processing pitch conversion, it comes to show an output sound signal in drawing 3 . The case where the pitch of such a sound signal is raised twice is considered. First, the spectrum information shown in drawing 2 as shown in drawing 4 is shifted so that it may become a twice as many frequency as this. Although the band of spectrum information spreads by 32kHz at this time, a band is deleted after carrying out the spreading band to to 16kHz of one half. Thereby, the number of data of a 0–16kHz band is set to 32 of one half from 64. If it changes into a time domain from a frequency domain in this condition, playback time amount will become short [4000 microseconds shown in drawing 3 to one half] to 2000 microseconds. In order to avoid this, data are interpolated to the spectrum information shown in drawing 4 , and as shown in drawing 5 , before shifting the number of data from 32, it increases to 64 of the same number. Interpolation of data is performed by the linear interpolation approach of adding the data of the midpoint between two data. Thus, after interpolating data and setting a measurement size to 64, inverse transformation is carried out to the data in a time domain from a frequency domain. Consequently, playback data serve as a sine wave with a frequency of 2kHz, while playback time amount has been 4000 microseconds, as shown in drawing 6 . That is, the pitch of sinusoidal data can be raised twice, without changing playback time amount.

[0027] Next, the case where the pitch of the sinusoidal data shown in drawing 2 is lowered to 1/2 is considered. In this case, as shown in drawing 7 to the spectrum information shown in drawing 2 , spectrum information is shifted so that it may become one half of frequencies. Thereby, the band of spectrum information narrows from 16kHz to 8kHz. If it changes into a time domain from a frequency domain in this condition, playback time amount will become long [to 8000 twice as many microseconds as this] from 4000 microseconds. In order to avoid this, data are operated on a curtailed schedule to the spectrum information shown in drawing 7 , and as shown in drawing 8 , before shifting the number of data from 64, it reduces to 32 (0–8kHz band) of the same number. Infanticide of data is performed by the approach of deleting the data of the midpoint between two data. Thus, after operating data on a curtailed schedule and setting a measurement size to 32, inverse transformation is carried out to the data in a time domain from a frequency domain. Consequently, playback data serve as a sine wave with a frequency of 0.5kHz, while playback time amount has been 4000 microseconds, as shown in drawing 9 . That is, the pitch of sinusoidal data can be lowered to one half, without changing playback time amount.

[0028] As explained above, in the pitch conversion in the above-mentioned operation gestalt, processing in the

accurate frequency domain where a noise is smaller than processing in a time domain Carry out using what is recorded as frequency data, such as MP3 and AAC, and in the process of the conversion to time amount from a frequency only by adding processing of the number step in software called a frequency shift, and a data interpolation/infanticide Making the pitch of playback voice adjustable at arbitration can be realized easily. Moreover, since the data of a frequency unit are outputted from the compression storage media on which compressed data, such as MP3 and AAC, was recorded, applying a burden to an arithmetic unit like a tape or CD by using this by big processing called data conversion from a time domain to a frequency domain is lost. Furthermore, since it has not carried out treating with the data of a time domain, the time stretch / compression which applied the previous operation gestalt to the degree which becomes without a jarring noise occurring in playback voice are explained.

[0029] Drawing 10 is drawing including the function of the time / pitch inverter concerning other operation gestalten of this invention showing the configuration of a voice data regenerative apparatus.

[0030] The storage media 31 to which a voice data regenerative apparatus outputs a compression sound signal in drawing 10 , The storage media I/F circuit 32 which receives the compression sound signal outputted from these storage media 31, DSP33 which has the function of the encoder 1 shown in drawing 1 , a decoder 2, and a time / pitch inverter in response to the output of the storage media I/F circuit 32 (digital signal processor), DAC34 which changes into an analog signal the digital signal outputted from DSP33, It has the clock speed adjustable circuit 35 which generates a clock signal in response to a clock speed setting signal, and the system clock generation circuit 36 which generates the clock signal of a system in response to the output of the clock speed adjustable circuit 35, and is constituted.

[0031] In such a configuration, since the read-out places of voice data are the storage media 31, a read-out rate becomes arbitrary, and if even the MIPS value (throughput per unit time amount) which decoding of read-out data takes is fulfilled, the system clock of DSP33 can be set up freely. Moreover, it has completed only with the configuration shown in drawing 10 , and if it is a system aiming only at audio playback, since it is not necessary to send the clock of the regular frequencies, such as a sampling frequency, to other circuits, the system clock of DAC34 can also be decided freely. That is, if there is no effect in a playback sound, it will not become a problem considering the system clock of the system shown in drawing 10 itself as adjustable. Moreover, it can perform making a system clock adjustable easily. Here, using this description, the pitch of voice data is beforehand changed by the approach of a previous operation gestalt, and the actuation which changes only playback time amount, without changing the pitch of a playback sound is explained by making adjustable the system clock of the whole system including DAC34. First, a time stretch is explained. In the system clock generation circuit 36, the system clock is beforehand set up so that it may be set to one half at the time of normal operation. It can perform making a system-wide clock adjustable easily with the device of a frequency divider etc. Moreover, although the MIPS value of DSP33 is reduced by half by setting a system clock to one half, it does not become a problem especially unless trouble is caused to decoding of input data. The hybrid filter bank 26 is operated to the data shown in drawing 2 and drawing 3 by the technique explained with the previous operation gestalt, and in case inverse transformation is carried out from a frequency domain to a time domain, the pitch of data is raised twice. Thereby, since the system clock given to DAC34 is 1/2 at the time of normal operation consequently, the pitch of the playback voice which inverse transformation was carried out and was obtained becomes the same as origin, as shown in drawing 11 , and playback time amount is extended twice.

[0032] On the other hand, it becomes the above-mentioned case and reverse, in the system clock generation circuit 36, the system clock is set up the twice at the time of normal operation beforehand, and in the case of time compression, the hybrid filter bank 26 is operated to the data shown in drawing 2 and drawing 3 by the technique explained with the previous operation gestalt, and in case inverse transformation of the data is carried out from a frequency domain to a time domain, the pitch of data is lowered to it 1/2. Thereby, since the system clock given to DAC34 is the twice at the time of normal operation consequently, the pitch of the playback voice which inverse transformation was carried out and was obtained becomes the same as origin, as shown in drawing 12 , and playback time amount is shortened by 1/2.

[0033] Thus, a time stretch / compression actuation can be realized easily, without reading like before in the case of a voice regeneration system including DAC34, and adding a speed regulating device, and big buffer memory and memory management equipment to it only by adding the adjustable circuit of the easy system clock for the configuration of a previous operation gestalt. that is , in the voice regeneration system which consist of an arithmetic circuit drive with the same system clock , and a DAC , it be possible to realize easily the time stretch / compression function of extend or shorten only playback time amount , only change the clock of operation in the configuration of the operation gestalt which mentioned the system clock above from the speed of arbitration for aim only at voice playback using the ability to be able to consider as adjustable , and fix the pitch of data .

[0034]

[Effect of the Invention] Since according to this invention interpolation/infanticide of data are performed and it was made to carry out inverse transformation to the voice data of time series data after shifting the spectrum of the voice data compressed as frequency data as explained above, the pitch of playback voice can be changed easily, without changing playback time amount. moreover, processing of the above-mentioned inverse transformation — in addition, since the frequency of the clock signal of operation at the time of changing a digitized voice signal into an analog sound signal was changed according to playback time amount, the playback time amount of playback voice can be extended / shortened easily, without changing a pitch.

[Translation done.]

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]

[Drawing 1] It is drawing showing the configuration of an MP3 encoder / decoder including the function of the time / pitch inverter concerning 1 operation gestalt of this invention.

[Drawing 2] It is drawing showing an example of the sinusoidal data in a frequency domain.

[Drawing 3] It is drawing showing the output sound signal corresponding to drawing 2 .

[Drawing 4] It is drawing showing the sinusoidal data which shifted the frequency of drawing 2 twice.

[Drawing 5] It is drawing showing the sinusoidal data which interpolated the data of drawing 4 .

[Drawing 6] It is drawing showing the output sound signal which carried out the pitch up of the sound signal of drawing 3 .

[Drawing 7] It is drawing showing the sinusoidal data which shifted the frequency of drawing 2 to 1/2.

[Drawing 8] It is drawing showing the sinusoidal data which thinned out and carried out the data of drawing 7 .

[Drawing 9] It is drawing showing the output sound signal which carried out the pitch down of the sound signal of drawing 3 .

[Drawing 10] It is drawing showing the voice playback structure of a system including the function of the time / pitch inverter concerning other operation gestalten of this invention.

[Drawing 11] It is drawing showing the output sound signal which carried out the time stretch of the sound signal of drawing 3 .

[Drawing 12] It is drawing showing the output sound signal which carried out time compression of the sound signal of drawing 3 .

[Drawing 13] It is drawing showing the 1 conventional example of pitch conversion of voice data.

[Drawing 14] It is drawing showing an example of cross fade processing.

[Drawing 15] It is drawing showing other conventional examples of pitch conversion of voice data.

[Drawing 16] It is drawing showing the 1 conventional technique of the time stretch of voice data.

[Description of Notations]

1 Encoder

2 Decoder

11 26 Hybrid filter bank

12 Mental Acoustic-Sense Analyzor

13 Repeat Loop Formation

14 Huffman Coding Section

15 Side Information Coding Section

16 Bit Stream Formation Section

21 Bit Stream Analysis Section 22
 Scale-Factor Decryption Section

23 Huffman Table Decryption Section

24 Huffman Coding Section

25 Reverse Quantization Section

111 SubBand Analysis Filter Bank

112 Adaptation Block Length MDCT

113,261 Clinch distorted reduction butterfly section

121,122 FFT

123 Non-Predictability Test Section

124 Mental Acoustic-Sense Entropy Evaluation Section

125 Signal Pair Mask Ratio Count Section

131 Nonlinear Quantization Section

132 Scale-Factor Count Section

133 Buffer Control Section

262 Reverse MDCT

263 SubBand Composition Filter Bank

[Translation done.]

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号
特開2001－356799
（P2001－356799A）
(43)公開日　平成13年12月26日(2001.12.26)

(51)Int.Cl.'	識別記号	ＦＩ		テーマコード'（参考）
Ｇ１０Ｌ　21/04		Ｇ１０Ｋ　15/04	３０２Ｄ	５Ｄ０４５
Ｇ１０Ｋ　15/04	３０２	Ｈ０３Ｍ　7/30	Ｚ	５Ｄ１０８
Ｇ１０Ｌ　13/00		Ｇ１０Ｌ　3/02	Ａ	５Ｊ０６４
Ｈ０３Ｍ　7/30			Ｃ	
		7/02	Ｄ	

審査請求　未請求　請求項の数５　ＯＬ　（全 13 頁）

(21)出願番号　特願2000－175065(P2000－175065)

(22)出願日　平成12年６月12日(2000.6.12)

(71)出願人　000003078
株式会社東芝
東京都港区芝浦一丁目１番１号

(72)発明者　岡崎　晶彦
東京都青梅市新町３丁目３番地の１　東芝デジタルメディアエンジニアリング株式会社内

(72)発明者　小島　能成
神奈川県川崎市幸区小向東芝町１番地　株式会社東芝マイクロエレクトロニクスセンター内

(74)代理人　100083806
弁理士　三好　秀和　（外７名）

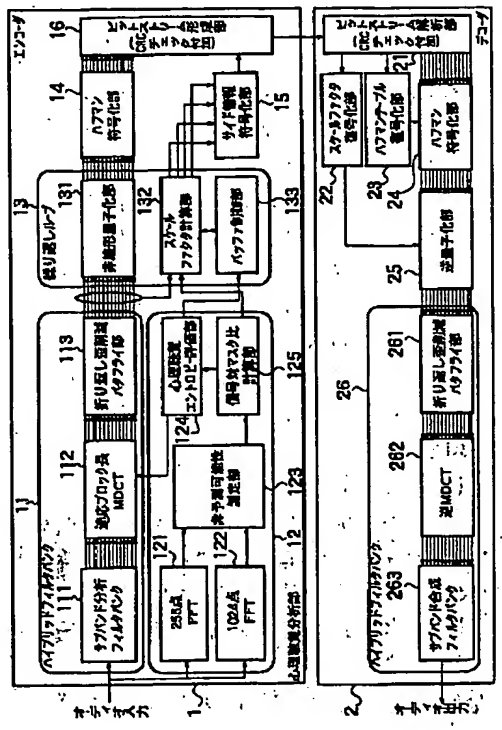
最終頁に続く

(54)【発明の名称】　タイム／ピッチ変換装置及びタイム／ピッチ変換方法

(57)【要約】

【課題】　この発明は、構成の大型化、処理の複雑化を招くことなく、かつ再生音質を損なうことなく再生音声のピッチ／再生時間を容易に変更できるタイム／ピッチ変換装置及びタイム／ピッチ変換方法を提供することを課題とする。

【解決手段】　この発明は、周波数データとして圧縮された音声データのスペクトルをシフトした後データの補間／間引きを行い、時系列データの音声データに逆変換するように構成される。



【特許請求の範囲】

【請求項１】　周波数データとして圧縮された音声データを入力し、周波数データとして圧縮された音声データを周波数領域から時間領域へ逆変換して時系列データの音声データを得る音声再生システムに具備されるタイム／ピッチ変換装置であって、

周波数データとして圧縮された音声データを周波数領域から時間領域へ逆変換して時系列データの音声データを得る際に、音声データのピッチ変換量に応じて周波数領域での音声データのスペクトルをシフトし、時系列データの音声データの再生周波数を決定するシフト手段と、

前記シフト手段によりシフトされて得られた周波数領域でのスペクトルに対して音声データを補間又は間引きし、シフト前後の周波数領域でのスペクトルの音声データ数を同一帯域幅で同一にする補間／間引き手段とを備え、

前記補間／間引き手段で得られた周波数領域での音声データが時系列データの音声データに逆変換される際に音声データのピッチを変えることを特徴とするタイム／ピッチ変換装置。

【請求項２】　周波数データとして圧縮された音声データを入力し、周波数データとして圧縮された音声データを周波数領域から時間領域へ逆変換して得られる時系列データのデジタル音声データをＤＡＣによりアナログ音声データに変換して再生する音声再生システムに具備されるタイム／ピッチ変換装置であって、

周波数データとして圧縮された音声データを周波数領域から時間領域へ逆変換して時系列データの音声データを得る際に、再生音声の再生時間に応じて周波数領域での音声データのスペクトルをシフトし、時系列データの音声データの再生周波数を決定するシフト手段と、

前記シフト手段によりシフトされて得られた周波数領域でのスペクトルに対して音声データを補間又は間引きし、シフト前後の周波数領域でのスペクトルの音声データ数を同一帯域幅で同一にする補間／間引き手段と、

再生音声の再生時間に応じて周波数が可変のクロック信号を生成し、生成したクロック信号を少なくとも前記ＤＡＣに供給するクロック生成手段とを備え、

前記クロック生成手段から供給されたクロック信号に基づいて前記ＤＡＣが時系列データのデジタル音声データをアナログ音声データに変換する際に音声データの再生時間を拡張／短縮することを特徴とするタイム／ピッチ変換装置。

【請求項３】　前記周波数データとして圧縮された音声データは、任意のデータ読み出し速度が可能なストレージメディアに格納されていることを特徴とする請求項１又は２記載のタイム／ピッチ変換装置。

【請求項４】　周波数データとして圧縮された音声データを入力し、

周波数データとして圧縮された音声データを周波数領域

から時間領域へ逆変換して時系列データの音声データを得る際に、音声データのピッチ変化量に応じて周波数領域での音声データのスペクトルをシフトし、時系列データの音声データの再生周波数を決定し、

シフトされて得られた周波数領域でのスペクトルに対して音声データを補間又は間引きし、シフト前後の周波数領域でのスペクトルの音声データ数を同一帯域幅で同一にし、

補間／間引きにより得られた周波数領域での音声データが時系列データの音声データに逆変換される際に音声データのピッチを変えることを特徴とするタイム／ピッチ変換方法。

【請求項５】　周波数データとして圧縮された音声データを入力し、

周波数データとして圧縮された音声データを周波数領域から時間領域へ逆変換して時系列データの音声データを得る際に、再生音声の再生時間に応じて周波数領域での音声データのスペクトルをシフトし、時系列データの音声データの再生周波数を決定し、

シフトされて得られた周波数領域でのスペクトルに対して音声データを補間又は間引きし、シフト前後の周波数領域でのスペクトルの音声データ数を同一帯域幅で同一にし、

再生音声の再生時間に応じて周波数が可変のクロック信号を生成し、生成したクロック信号を少なくともＤＡＣに供給し、

周波数領域から時間領域への逆変換で得られた時系列データのデジタル音声データを、前記ＤＡＣが供給されたクロック信号に基づいてアナログ音声データに変換する際に音声データの再生時間を拡張／短縮することを特徴とするタイム／ピッチ変換方法。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、入力が時系列データではなく、周波数データである信号を再生するシステムにおいて、再生音声のタイム又はピッチ変換を行うタイム／ピッチ変換装置及びタイム／ピッチ変換方法に関する。

【０００２】

【従来の技術】レコーディングのピッチ変換用エフェクタ、コマーシャル制作などの演奏時間を変更する装置、会議録音、インタビュー、ニュースなどの話速変換装置、カラオケなどのピッチコントローラなどの様々な用途でピッチ変換の技術が必要となっている。

【０００３】従来、音声データのピッチを変換する手法としては、時間領域での処理と周波数領域での処理の２通りに大別される。時間領域での処理では、時間軸上で波形の不連続点が発生し、音声再生時に耳障りなノイズとして現れていた。これに比べて周波数領域での処理では、このような不連続点の発生がないのでノイズを生成

3

することはなかった。しかし、録音テープやＣＤなどのメディアでは、音声が時系列データとして記録されているため、周波数領域においてピッチ変換を行なうためには、ＦＦＴ（高速フーリエ変換）などの時間←→周波数変換を行なう必要があった。しかし、ＦＦＴを行なうには多くの演算を行なわなければならず、演算回路の処理能力が大きくなければならないといった欠点があった。

【０００４】次に、ピッチ変換について詳細に説明する。

【０００５】ピッチを変換する手法としては、上述したように、

（ａ）時間領域でのデータ処理によるもの

（ｂ）周波数領域でのデータ処理によるもの

の２通りに大別されるが、主にカラオケのキーコントロールなどの簡易的なシステムには前者の手法が、楽器などの音質に関する要求が厳しいシステムには後者の手法が用いられていた。

【０００６】図１３に上記（ａ）の手法によるピッチ変換の一例を示す。時間領域での処理では、時系列データの再生速度を制御することでピッチのアップ／ダウンを行うが、図１３に示すように、同時に再生時間が短縮あるいは延長されていることに注意が必要である。すなわち、ピッチを下げた場合には同時に再生時間が延長され、一方ピッチを上げた場合には同時に再生時間が短縮される。ここでは、再生時間は変えず、ピッチのみを変換することを目的としており、再生時間は元データのそれと同じでなければならない。そのため、元データのピッチを下げた場合には、必ずどこかで重複部分が生じ、またピッチを上げた場合に必ずどこかでデータの欠落部分が生じてしまう。これらは、時系列上でのデータの不連続となるので、そのまま再生するとノイズが発生し音質が悪くなってしまう。このような不具合を回避するための技術として、クロスフェード処理がある。この処理は、図１４に示すように、ピッチを下げた場合は連続波形の終了をフェードアウトし、それと同時に次の連続波形の開始をフェードインしてクロスフェード連続を行なう。これによって接続点でのノイズは減少する。一方、ピッチを上げた場合には、データの欠落部分を補うために同じデータを２回再生し、同様にクロスフェード連続によって接続点でのノイズは減少する。しかし、このクロスフェード処理では、フェードアウト音とフェードイン音の位相が逆転している場合などは良い結果を得ることができないこともある。また、再生音に周期的なうねりが発生することも問題視されていた。

【０００７】次に、上記（ｂ）の処理でピッチを変化する手法は、図１５に示すように周波数軸上でデータをシフトすることで容易にピッチ変化を行なうことができ、また時間軸上での不連続点も発生しない。このため、上記（ａ）に比べて再生音の音質が良いのが特徴である。しかしながら、テープやＣＤ等から出力される音声データは時系列データであり、これを時間領域から周波数領域に変換するためには、ＦＦＴなどの演算処理が必要である。この演算処理は、主に演算回路とメモリから構成されるＤＳＰ（デジタル・シグナル・プロセッサ）などの装置またはシステムで行なうことができるが、多くの演算を行なわなければならず、演算回路の処理能力が大きくなくてはならないといった欠点があった。

4

【０００８】次に、音声データの再生時間を変えるタイム変換技術について説明する。

【０００９】再生音のピッチを変えることなく再生時間の短縮、延長のみを行うことをタイムストレッチ／コンプレッションといい、主に話速変換やサンプラーという機器に用いられている。これは、上述したピッチ変換の技術を応用して実現できる。

【００１０】再生速度を遅くして再生時間を長くした場合は、前述した理由から再生音のピッチが下がってしまうので、これをピッチ変換の技術を使って元のピッチに戻すように操作する。これにより、図１６に示すようにピッチはそのままで再生時間のみを延長することができる。一方、再生時間を短縮するにはこれとは逆の操作を行えばよい。

【００１１】これまでよく利用されてきたＣＤ、音楽テープなどの時系列データをそのまま記録したメディアを再生し、タイムストレッチ／コンプレッションを行う場合には、再生速度をコントロールする装置を使ってメディアからの読み出し速度を可変にさせるか、あるいは再生速度はそのままでシステムに大きなバッファメモリを持たせて再生時間の調節を行うような手法が採用されていた。ただし、両者とも複雑な付加装置や大掛かりな処理が必要となり、簡単に実現できるまでは至らなかった。

【００１２】

【発明が解決しようとする課題】以上説明したように、音声データのピッチを変換する従来の変換手法の内、時間領域での処理においては、音声データの不連続を回避するためのクロスフェード処理を行っているが、この処理を行っても再生音からノイズを確実に除去することは困難であり、音質が劣化するといった不具合を招いていた。一方、周波数領域での処理においては、音声データを時間領域から周波数領域へ変換する処理が必要となり、この処理を行うためには、大規模な構成と多大な時間が必要になるといった不具合を招いていた。

【００１３】そこで、本発明は、上記に鑑みてなされたものであり、その目的とするところは、構成の大型化、処理の複雑化を招くことなく、かつ再生音質を損なうことなく再生音声のピッチ／再生時間を容易に変更できるタイム／ピッチ変換装置及びタイム／ピッチ変換方法を提供することにある。

【００１４】

【課題を解決するための手段】上記目的を達成するため

5

に、課題を解決する第１の手段は、周波数データとして圧縮された音声データを入力し、周波数データとして圧縮された音声データを周波数領域から時間領域へ逆変換して時系列データの音声データを得る音声再生システムに具備されるタイム／ピッチ変換装置であって、周波数データとして圧縮された音声データを周波数領域から時間領域へ逆変換して時系列データの音声データを得る際に、音声データのピッチ変換量に応じて周波数領域での音声データのスペクトルをシフトし、時系列データの音声データの再生周波数を決定するシフト手段と、前記シフト手段によりシフトされて得られた周波数領域でのスペクトルに対して音声データを補間又は間引きし、シフト前後の周波数領域でのスペクトルの音声データ数を同一帯域幅で同一にする補間／間引き手段とを備え、前記補間／間引き手段で得られた周波数領域での音声データが時系列データの音声データに逆変換される際に音声データのピッチを変えることを特徴とする。

【００１５】第２の手段は、周波数データとして圧縮された音声データを入力し、周波数データとして圧縮された音声データを周波数領域から時間領域へ逆変換して得られる時系列データのデジタル音声データをＤＡＣによりアナログ音声データに変換して再生する音声再生システムに具備されるタイム／ピッチ変換装置であって、周波数データとして圧縮された音声データを周波数領域から時間領域へ逆変換して時系列データの音声データを得る際に、再生音声の再生時間に応じて周波数領域での音声データのスペクトルをシフトし、時系列データの音声データの再生周波数を決定するシフト手段と、前記シフト手段によりシフトされて得られた周波数領域でのスペクトルに対して音声データを補間又は間引きし、シフト前後の周波数領域でのスペクトルの音声データ数を同一帯域幅で同一にする補間／間引き手段と、再生音声の再生時間に応じて周波数が可変のクロック信号を生成し、生成したクロック信号を少なくとも前記ＤＡＣに供給するクロック生成手段とを備え、前記クロック生成手段から供給されたクロック信号に基づいて前記ＤＡＣが時系列データのデジタル音声データをアナログ音声データに変換する際に音声データの再生時間を拡張／短縮することを特徴とする。

【００１６】第３の手段は、前記第１又は第２の手段において、前記周波数データとして圧縮された音声データは、任意のデータ読み出し速度が可能なストレージメディアに格納されていることを特徴とする。

【００１７】第４の手段は、周波数データとして圧縮された音声データを入力し、周波数データとして圧縮された音声データを周波数領域から時間領域へ逆変換して時系列データの音声データを得る際に、音声データのピッチ変化量に応じて周波数領域での音声データのスペクトルをシフトし、時系列データの音声データの再生周波数を決定し、シフトされて得られた周波数領域でのスペクトルに対して音声データを補間又は間引きし、シフト前後の周波数領域でのスペクトルの音声データ数を同一帯域幅で同一にし、補間／間引きにより得られた周波数領域での音声データが時系列データの音声データに逆変換される際に音声データのピッチを変えることを特徴とする。

6

【００１８】第５の手段は、周波数データとして圧縮された音声データを入力し、周波数データとして圧縮された音声データを周波数領域から時間領域へ逆変換して時系列データの音声データを得る際に、再生音声の再生時間に応じて周波数領域での音声データのスペクトルをシフトし、時系列データの音声データの再生周波数を決定し、シフトされて得られた周波数領域でのスペクトルに対して音声データを補間又は間引きし、シフト前後の周波数領域でのスペクトルの音声データ数を同一帯域幅で同一にし、再生音声の再生時間に応じて周波数が可変のクロック信号を生成し、生成したクロック信号を少なくともＤＡＣに供給し、周波数領域から時間領域への逆変換で得られた時系列データのデジタル音声データを、前記ＤＡＣが供給されたクロック信号に基づいてアナログ音声データに変換する際に音声データの再生時間を拡張／短縮することを特徴とする。

【００１９】

【発明の実施の形態】以下、図面を用いて本発明の一実施形態を説明する。

【００２０】図１は本発明の一実施形態に係るタイム／ピッチ変換装置の機能を含む、ＭＰ３エンコーダ／デコーダの構成を示す。

【００２１】この実施形態では、ＭＰＥＧ音声圧縮方式の一つであるＭＰ３方式により圧縮された圧縮音声を再生する際のピッチ変換について説明する。なお、音声データが周波数データであれば全て適用可能であるので、ＭＰ３の他にＡＣＣ等のＭＰＥＧ音声圧縮方式であっても実施可能であり、また音声圧縮は特にＭＰＥＧ方式に限定されることはない。ＭＰＥＧによる圧縮音声データは、すでに周波数データとして記録されているので、時系列データを記録したメディアの再生のように周波数・時間変換する必要はない。この点を利用し、さらにＭＰＥＧの圧縮音声データのデコード時に行われるフィルタ演算処理をほとんど変更することなく、フィルタ演算処理のアルゴリズムを実行するソフトウェアに数ステップのプログラムを追加するだけで周波数領域でのスペクトル情報の操作を行い、再生音声のピッチ変換を容易に実現するようにしている。

【００２２】図１において、この実施形態のＭＰ３エンコーダ／デコーダは、時系列データである音声データを入力し、この音声データを従来から知られているＭＰ３の圧縮方式により周波数領域でのデータに圧縮変換するエンコーダ１と、このエンコーダ１の周波数領域での出力を受けて、この出力を時系列データに逆変換して時系

7

列データの音声データとして出力するデコーダ２を備えて構成されている。エンコーダ１は、ハイブリッドフィルタバンク１１と、心理聴覚分析部１２と、繰り返しループ１３と、繰り返しループ１３の出力を受けてハフマン符号化処理を行うハフマン符号化部１４と、繰り返しループ１３の出力を受けてサイド情報の符号化を行うサイド情報符号化部１５と、ハフマン符号化部１４の出力とサイド情報符号化部１５の出力を受けてビットストリームを形成するビットストリーム形成部１６を備え、ハイブリッドフィルタバンク１１は、サブバンド分析フィルタバンク１１１と、適応ブロック長MDCT１１２と、折り返し歪削減バタフライ部１１３を備え、心理聴覚分析部１２は、２５６点のFET（高速フーリエ変換）１２１と、１０２４点のＦＦＴ１２２と、被予測可能性測定部１２３と、心理聴覚エントロピー評価部１２４と、信号対マスク比計算部１２５を備え、繰り返しループ１３は、非線形量子化部１３１と、スケールファクタ計算部１３２と、バッファ制御部１３３を備えて構成されている。

【００２３】デコーダ２は、エンコーダ１のビットストリーム形成部１６の周波数領域での出力を受けて、ビットストリームを解析するビットストリーム解析部２１と、ビットストリーム解析部２１の出力を受けて、スケールファクタ復号化を行うスケールファクタ復号化部２２と、ビットストリーム解析部２１の出力を受けて、ハフマンテーブル復号化を行うハフマンテーブル復号化部２３と、ビットストリーム解析部２１及びハフマンテーブル復号化部２３の出力を受けて、ハフマン符号化を行うハフマン符号化部２４と、スケールファクタ復号化部２２及びハフマン符号化部２４の出力を受けて、逆量子化を行いスペクトル情報を得る逆量子化部２５と、逆量子化部２５の出力を受けて、時系列データとしての音声データを再生し、かつこの再生過程においてこの実施形態の特徴となるピッチ変換処理を行うシフト手段と補間／間引き手段を含むハイブリッドフィルタバンク２６を備え、ハイブリッドフィルタバンク２６は、逆量子化部２５で得られたスペクトル情報をバタフライ演算する折り返し歪削減バタフライ部２６１と、折り返し歪削減バタフライ部２６１の出力を受けて、逆フーリエ変換を行う逆MDCT２６２と、逆MDCT２６２の出力を受けて、サブバンド合成を行うサブバンド合成フィルタバンク２６３を備えて構成される。

【００２４】デコーダ２のハイブリッドフィルタバンク２６では、バタフライ演算、逆MDCT、QMF合成の処理を行っているが、これらの処理はソフトウェアによる１つのまとまったアルゴリズムとして処理される。また、このアルゴリズムでは、ピッチ変換処理を行うために、シフト手段によりまず周波数・時間変換を行う際に周波数領域でのスペクトル情報のシフトを行い、再生音声の周波数を決定し、補間／間引き手段によりシフトしたスペクトル情報に対して周波数領域でのデータの補間又は間引きの処理を行い、データの個数をそろえる。これにより、ピッチを変更すると共に、スペクトル情報を時間領域に戻した場合に、再生時間が変わらないようにする。

8

【００２５】次に、上記処理について、図２に示すような周波数領域の正弦波データを一例として、図３～図９を参照して説明する。以下、ＦＦＴ／逆ＦＦＴを用いて帯域０～１６ｋＨｚのスペクトル情報についてシミュレーションした結果に基づいて説明する。逆ＦＦＴに入力するデータは、１ｋＨｚの正弦波、サンプリング周波数＝３２ｋＨｚ、サンプル数＝６４とする。

【００２６】ピッチ変換の処理をしない場合には、出力音声信号は図３に示すようになる。このような音声信号のピッチを２倍に上げる場合を考える。まず、図４に示すように図２に示すスペクトル情報を２倍の周波数になるようにシフトする。このとき、スペクトル情報の帯域は３２ｋＨｚまでに広がるが、広がった帯域を半分の１６ｋＨｚまでとして以降の帯域を削除する。これにより、０～１６ｋＨｚの帯域のデータ数は６４から半分の３２となる。この状態で周波数領域から時間領域に変換すると、再生時間が図３に示す４０００μｓから半分の２０００μｓに短くなってしまう。これを回避するために、図４に示すスペクトル情報に対してデータを補間し、図５に示すようにデータ数を３２からシフトする前と同数の６４に増やす。データの補間は、例えば２つのデータ間の中間点のデータを加える一次補間方法によって行われる。このようにして、データを補間してサンプル数を６４にした後、周波数領域から時間領域でのデータに逆変換する。その結果、再生データは、図６に示すように再生時間が４０００μｓのままで周波数２ｋＨｚの正弦波となる。すなわち、再生時間を変えることなく、正弦波データのピッチを２倍に上げることができる。

【００２７】次に、図２に示す正弦波データのピッチを１／２倍に下げる場合を考える。この場合には、図２に示すスペクトル情報に対して図７に示すようにスペクトル情報を１／２の周波数となるようにシフトする。これにより、スペクトル情報の帯域は１６ｋＨｚから８ｋＨｚに狭まる。この状態で周波数領域から時間領域に変換すると、再生時間が４０００μｓから２倍の８０００μｓに長くなってしまう。これを回避するために、図７に示すスペクトル情報に対してデータを間引きし、図８に示すようにデータ数を６４からシフトする前と同数の３２（０～８ｋＨｚの帯域）に減らす。データの間引きは、例えば２つのデータ間の中間点のデータを削除する方法によって行われる。このようにして、データを間引きしてサンプル数を３２にした後、周波数領域から時間領域でのデータに逆変換する。その結果、再生データは、図９に示すように再生時間が４００００μｓのままで

周波数０．５ｋＨｚの正弦波となる。すなわち、再生時間を変えることなく、正弦波データのピッチを１／２に下げることができる。

【００２８】以上説明したように、上記実施形態におけるピッチ変換において、時間領域での処理よりもノイズが小さく精度が良い周波数領域での処理を、ＭＰ３、ＡＡＣなどの周波数データとして記録されているものを用いて行い、周波数から時間への変換の過程において、周波数シフト、データ補間／間引きというソフトウェアにおける数ステップの処理を加えるだけで、再生音声のピッチを任意に可変とすることが容易に実現できる。また、ＭＰ３、ＡＡＣ等の圧縮データが記録された圧縮ストレージメディアからは周波数単位のデータが出力されるので、これを利用することで、テープやＣＤ等のように時間領域から周波数領域へのデータ変換といった大きな処理で演算装置に負担をかけることがなくなる。さらに、時間領域のデータのまま扱うことをしていないので、再生音声に耳障りなノイズが発生することもなくなる次に、先の実施形態を応用したタイムストレッチ／コンプレッションについて説明する。

【００２９】図１０はこの発明の他の実施形態に係るタイム／ピッチ変換装置の機能を含む、音声データ再生装置の構成を示す図である。

【００３０】図１０において、音声データ再生装置は、圧縮音声信号を出力するストレージメディア３１と、このストレージメディア３１から出力された圧縮音声信号を受けるストレージメディアＩ／Ｆ回路３２と、ストレージメディアＩ／Ｆ回路３２の出力を受けて、図１に示すエンコーダ１とデコーダ２ならびにタイム／ピッチ変換装置の機能を有するＤＳＰ（デジタル・シグナル・プロセッサ）３３と、ＤＳＰ３３から出力されるデジタル信号をアナログ信号に変換するＤＡＣ３４と、クロックスピード設定信号を受けてクロック信号を生成するクロックスピード可変回路３５と、クロックスピード可変回路３５の出力を受けてシステムのクロック信号を生成するシステムクロック生成回路３６とを備えて構成される。

【００３１】このような構成において、音声データの読み出し先がストレージメディア３１であるため読み出し速度が任意となり、読み出しデータのデコードに要するＭＩＰＳ値（単位時間あたりの処理能力）さえ満たしていれば、ＤＳＰ３３のシステムクロックを自由に設定することができる。また、図１０に示す構成のみで完結しており、音声の再生だけを目的としたシステムであれば、他の回路にサンプリング周波数などの決まった周波数のクロックを送る必要がないので、ＤＡＣ３４のシステムクロックも自由に決めることができる。すなわち、再生音に影響がなければ、図１０に示すシステムのシステムクロックそのものを可変としても問題とはならない。また、システムクロックを可変とすることは容易に行える。ここでは、この特徴を利用して、先の実施形態の方法で音声データのピッチをあらかじめ変えておき、ＤＡＣ３４を含めたシステム全体のシステムクロックを可変とすることで、再生音のピッチを変えずに再生時間のみを変える動作を説明する。まず、タイムストレッチについて説明する。システムクロック生成回路３６において、システムクロックを通常動作時の１／２になるようにあらかじめ設定しておく。システム全体のクロックを可変とすることは分周回路の工夫などで簡単に行うことができる。また、システムクロックを１／２にすることでＤＳＰ３３のＭＩＰＳ値は半減するが、入力データのデコードに支障をきたさないかぎり特に問題になることはない。先の実施形態で説明した手法で図２ならびに図３に示すデータに対してハイブリッドフィルタバンク２６を操作し、周波数領域から時間領域へ逆変換する際にデータのピッチを２倍に上げる。これにより、ＤＡＣ３４に与えられるシステムクロックは通常動作時の１／２であるので、その結果、逆変換されて得られた再生音声のピッチは、図１１に示すように元と同じになり、かつ再生時間が２倍に拡張される。

【００３２】一方、タイムコンプレッションの場合には、上記の場合と逆になり、システムクロック生成回路３６において、あらかじめシステムクロックを通常動作時の２倍に設定しておき、先の実施形態で説明した手法で図２ならびに図３に示すデータに対してハイブリッドフィルタバンク２６を操作し、データを周波数領域から時間領域へ逆変換する際にデータのピッチを１／２倍に下げる。これにより、ＤＡＣ３４に与えられるシステムクロックは通常動作時の２倍であるので、その結果、逆変換されて得られた再生音声のピッチは、図１２に示すように元と同じになり、かつ再生時間が１／２倍に短縮される。

【００３３】このように、ＤＡＣ３４を含めた音声再生システムの場合に、先の実施形態の構成に簡単なシステムクロックの可変回路を加えるだけで、従来のように読み出し速度制御装置や大きなバッファメモリ及びメモリマネージメント装置を付加することなく、タイムストレッチ／コンプレッション操作が容易に実現できる。すなわち、同一のシステムクロックで駆動される演算回路とＤＡＣから構成される音声再生システムでは、音声再生のみを目的とすることでシステムクロックを任意のスピードに可変とすることができることを利用して、前述した実施形態の構成における動作クロックを変化させるだけで、データのピッチを固定したまま再生時間のみを延長又は短縮するタイムストレッチ／コンプレッション機能を容易に実現することが可能である。

【００３４】

【発明の効果】以上説明したように、この発明によれば、周波数データとして圧縮された音声データのスペクトルをシフトした後データの補間／間引きを行い、時系

列データの音声データに逆変換するようにしたので、再生時間を変えることなく再生音声のピッチを容易に変更することができる。また、上記逆変換の処理に加えて、デジタル音声信号をアナログ音声信号に変換する際の動作クロック信号の周波数を再生時間に応じて変えるようにしたので、ピッチを変えることなく再生音声の再生時間を容易に拡張／短縮することができる。

【図面の簡単な説明】

【図１】この発明の一実施形態に係るタイム／ピッチ変換装置の機能を含むＭＰ３エンコーダ／デコーダの構成を示す図である。

【図２】周波数領域での正弦波データの一例を示す図である。

【図３】図２に対応した出力音声信号を示す図である。

【図４】図２の周波数を２倍にシフトした正弦波データを示す図である。

【図５】図４のデータを補間した正弦波データを示す図である。

【図６】図３の音声信号をピッチアップした出力音声信号を示す図である。

【図７】図２の周波数を１／２倍にシフトした正弦波データを示す図である。

【図８】図７のデータを間引きした正弦波データを示す図である。

【図９】図３の音声信号をピッチダウンした出力音声信号を示す図である。

【図１０】この発明の他の実施形態に係るタイム／ピッチ変換装置の機能を含む音声再生システムの構成を示す図である。

【図１１】図３の音声信号をタイムストレッチした出力音声信号を示す図である。

【図１２】図３の音声信号をタイムコンプレッションした出力音声信号を示す図である。

【図１３】音声データのピッチ変換の一従来例を示す図である。

【図１４】クロスフェード処理の一例を示す図である。

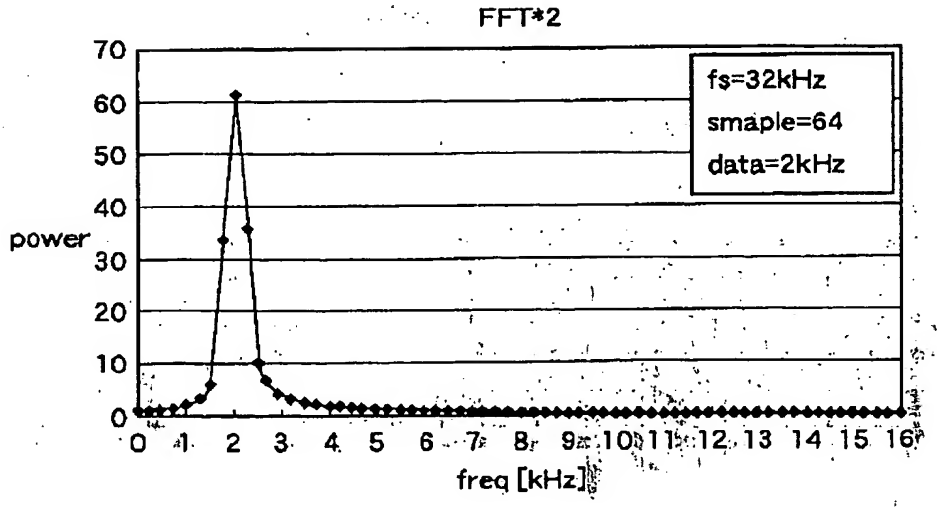
【図１５】音声データのピッチ変換の他の従来例を示す図である。

【図１６】音声データのタイムストレッチの一従来手法を示す図である。

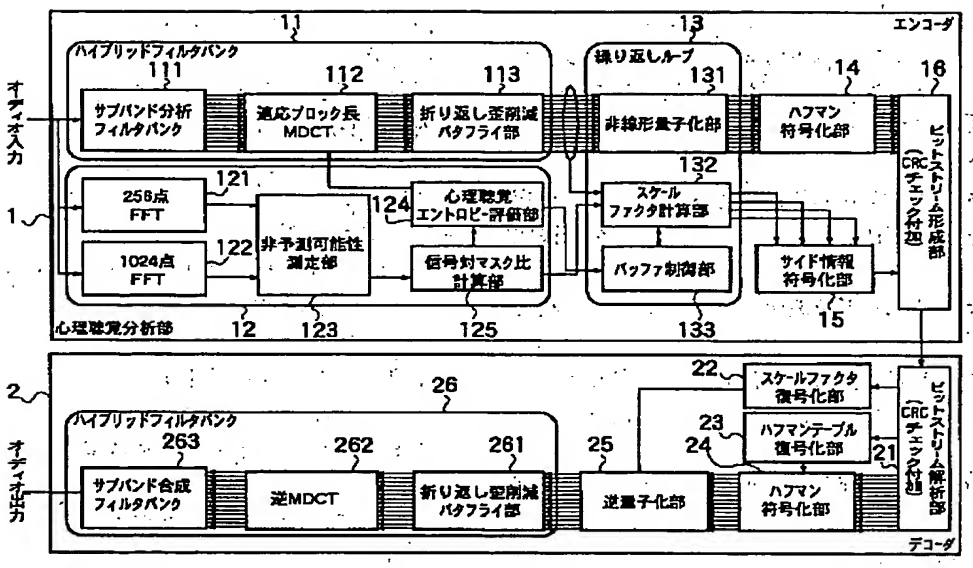
【符号の説明】

１　エンコーダ
２　デコーダ
１１，２６　ハイブリッドフィルタバンク
１２　心理聴覚分析部
１３　繰り返しループ
１４　ハフマン符号化部
１５　サイド情報符号化部
１６　ビットストリーム形成部
２１　ビットストリーム解析部
２２　スケールファクタ復号化部
２３　ハフマンテーブル復号化部
２４　ハフマン符号化部
２５　逆量子化部
１１１　サブバンド分析フィルタバンク
１１２　適応ブロック長ＭＤＣＴ
１１３，２６１　折り返し歪削減バタフライ部
１２１，１２２　ＦＦＴ
１２３　非予測可能性測定部
１２４　心理聴覚エントロピー評価部
１２５　信号対マスク比計算部
１３１　非線形量子化部
１３２　スケールファクタ計算部
１３３　バッファ制御部
２６２　逆ＭＤＣＴ
２６３　サブバンド合成フィルタバンク

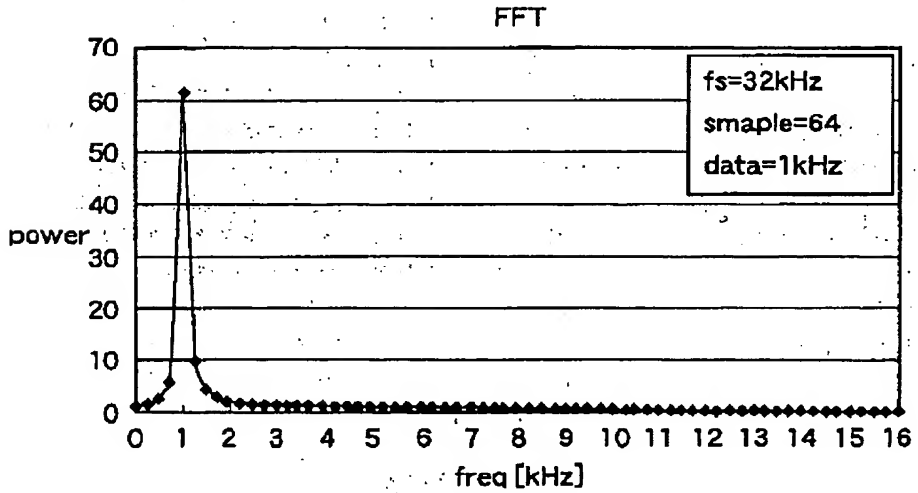
【図５】



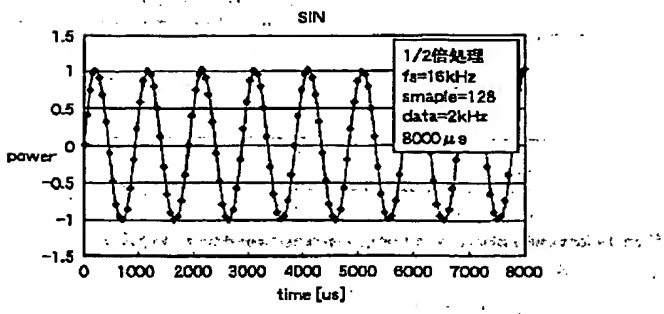
【図１】


エンコーダ
ハイブリッドフィルタバンク
サブバンド分析フィルタバンク
適応ブロック長MDCT
折り返し歪削減バタフライ部
繰り返しループ
非線形量子化部
ハフマン符号化部
ビットストリーム形成部（CRCチェック付加）
オーディオ入力
256点FFT
1024点FFT
非予測可能性測定部
心理聴覚エントロピー評価部
信号対マスク比計算部
スケールファクタ計算部
バッファ制御部
サイド情報符号化部
心理聴覚分析部
スケールファクタ復号化部
ハフマンテーブル復号化部
ビットストリーム解析部（CRCチェック付加）
サブバンド合成フィルタバンク
逆MDCT
逆量子化部
ハフマン符号化部
オーディオ出力
デコーダ

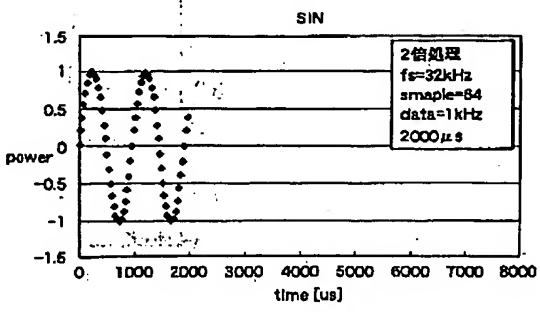
【図２】


FFT
fs=32kHz
smaple=64
data=1kHz
power
freq [kHz]

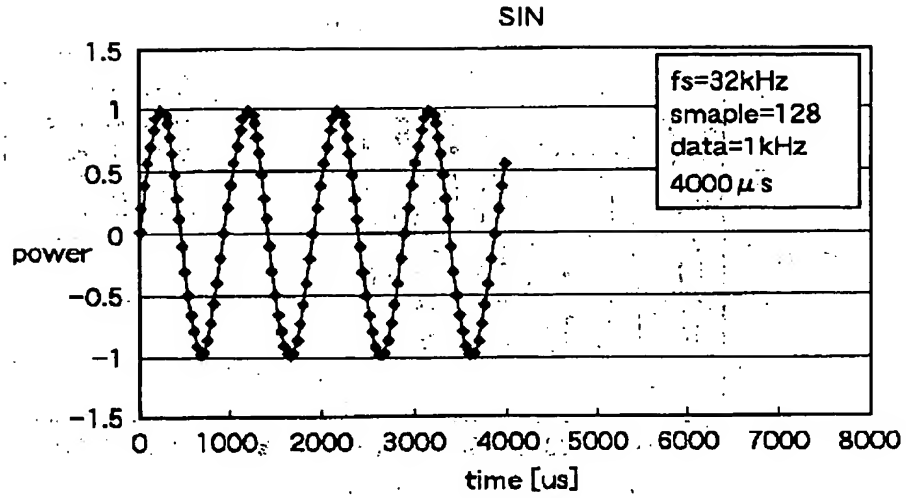
【図１１】


SIN
1/2倍処理
fs=16kHz
smaple=128
data=2kHz
8000μs
power
time [us]

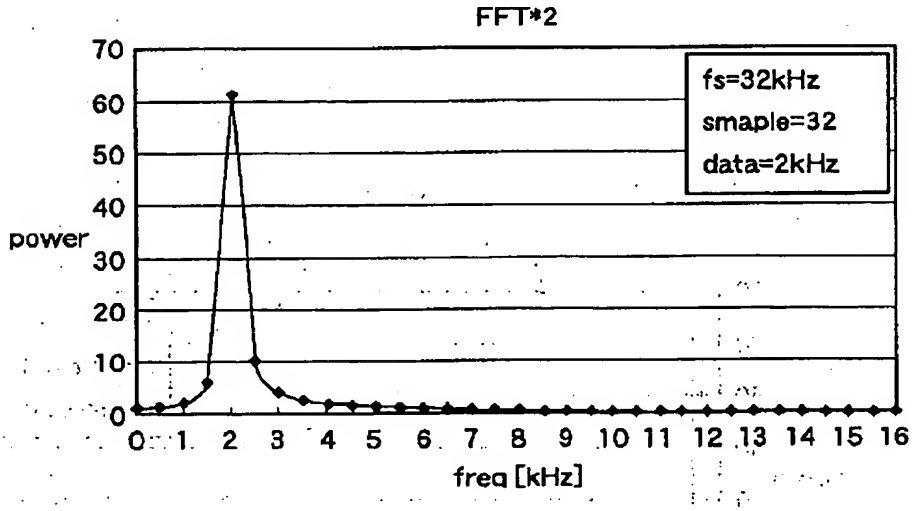
【図１２】


SIN
2倍処理
fs=32kHz
smaple=64
data=1kHz
2000μs
power
time [us]

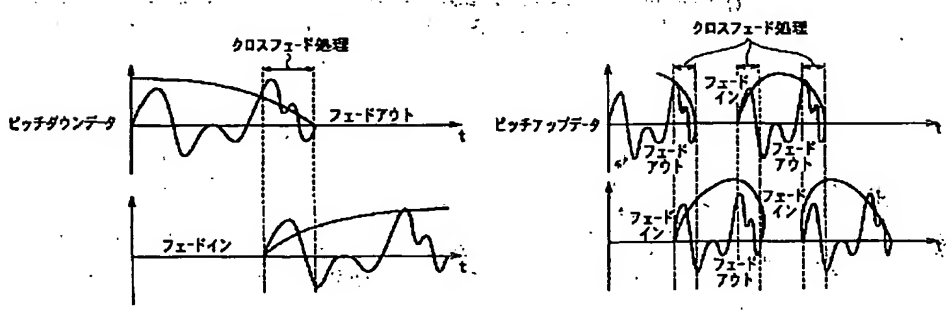
【図３】


SIN
fs=32kHz
smaple=128
data=1kHz
4000μs
power
time [us]

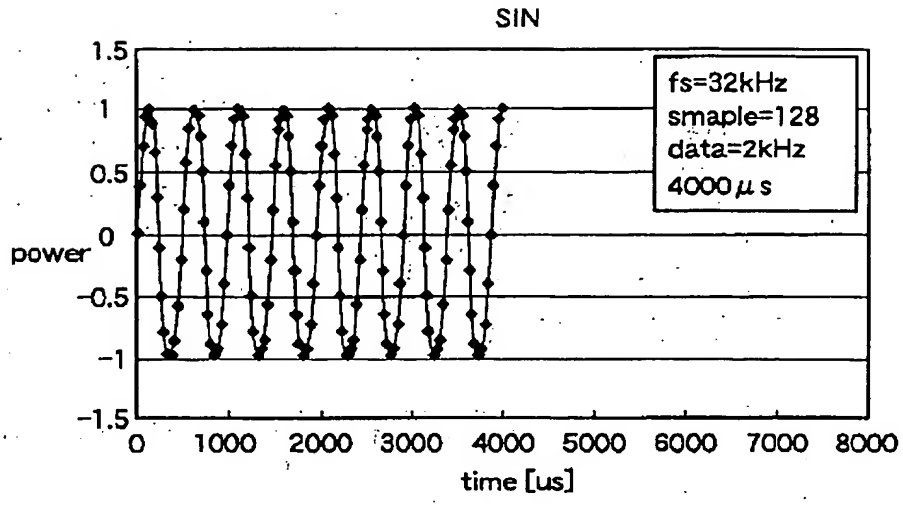
【図４】


FFT*2
fs=32kHz
smaple=32
data=2kHz
power
freq [kHz]

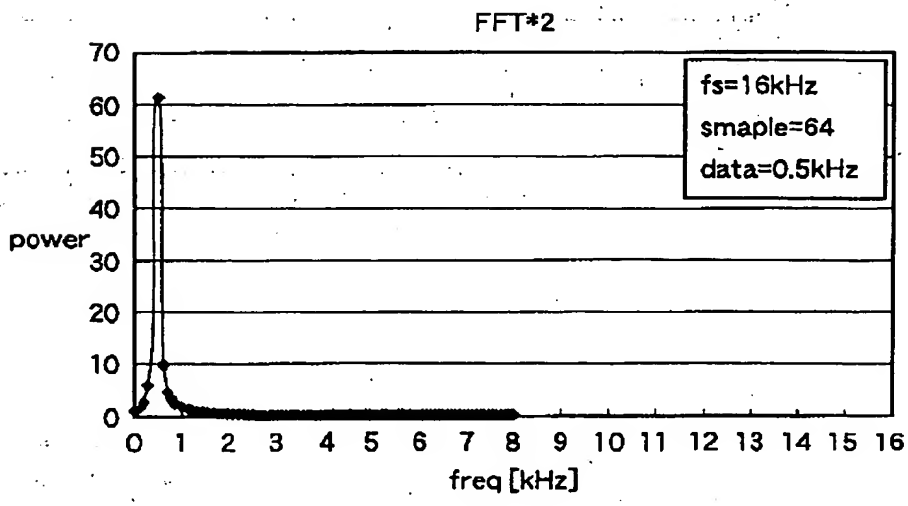
【図１４】


クロスフェード処理
ピッチダウンデータ
フェードアウト
フェードイン
クロスフェード処理
ピッチアップデータ
フェードイン
フェードアウト
フェードアウト
フェードイン
フェードイン
フェードアウト

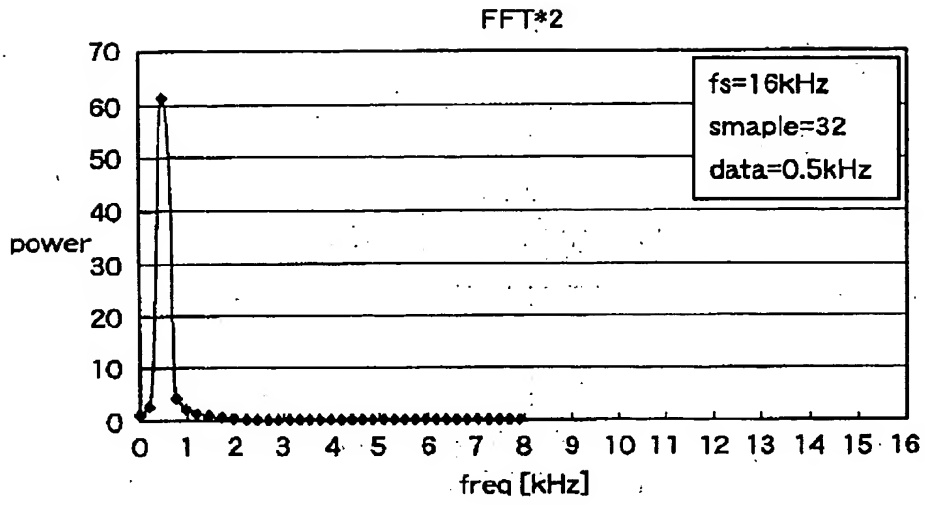
【図6】


SIN
fs=32kHz
smaple=128
data=2kHz
4000μs
power
1.5
1
0.5
0
-0.5
-1
-1.5
0
1000
2000
3000
4000
5000
6000
7000
8000
time [us]

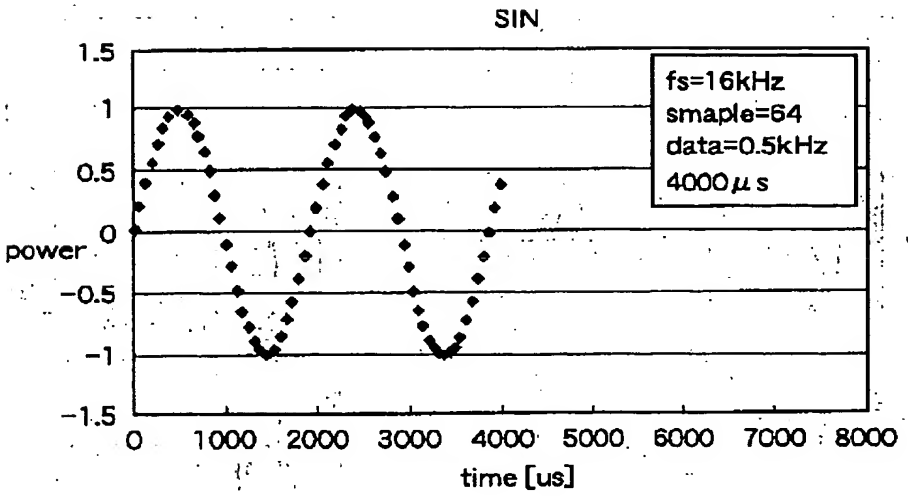
【図7】


FFT*2
fs=16kHz
smaple=64
data=0.5kHz
power
70
60
50
40
30
20
10
0
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
freq [kHz]

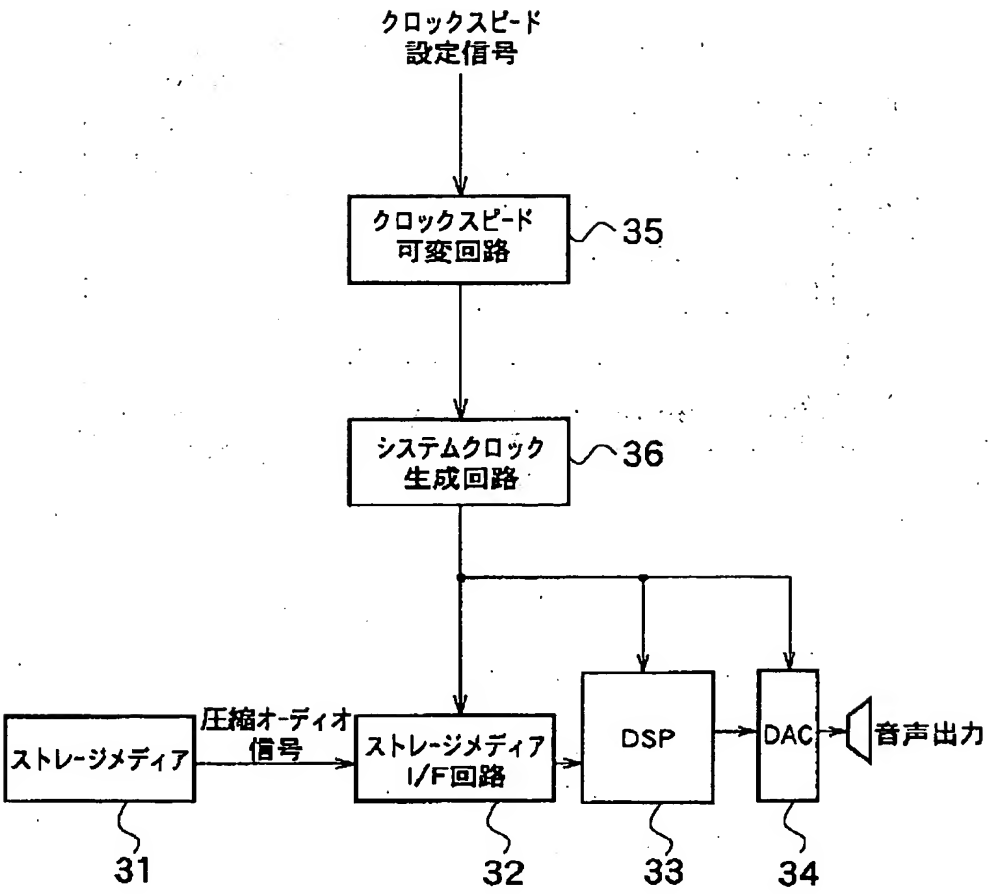
【図８】


FFT*2
fs=16kHz
smaple=32
data=0.5kHz
power
freq [kHz]
0
10
20
30
40
50
60
70
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

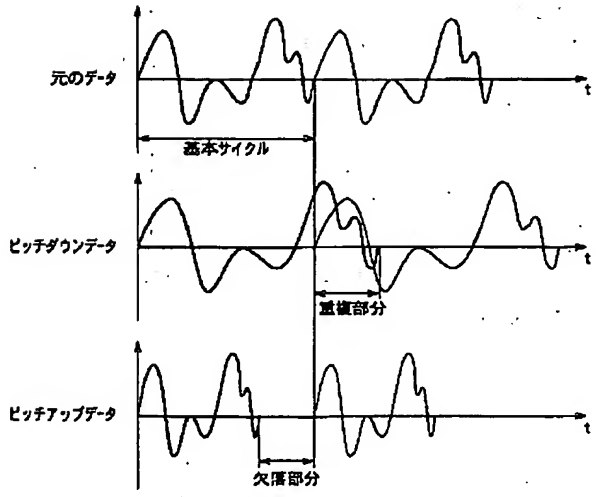
【図９】


SIN
fs=16kHz
smaple=64
data=0.5kHz
4000μs
power
time [us]
1.5
1
0.5
0
-0.5
-1
-1.5
0 1000 2000 3000 4000 5000 6000 7000 8000

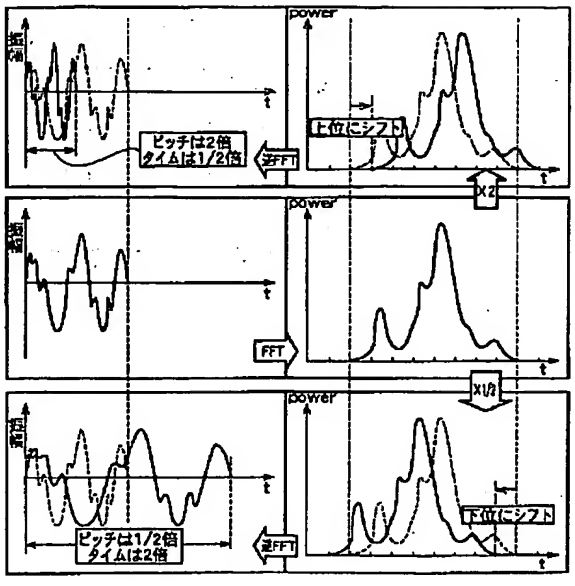
【図１０】


クロックスピード
設定信号
クロックスピード
可変回路
35
システムクロック
生成回路
36
ストレージメディア
圧縮オーディオ
信号
ストレージメディア
I/F回路
DSP
DAC
音声出力
31
32
33
34

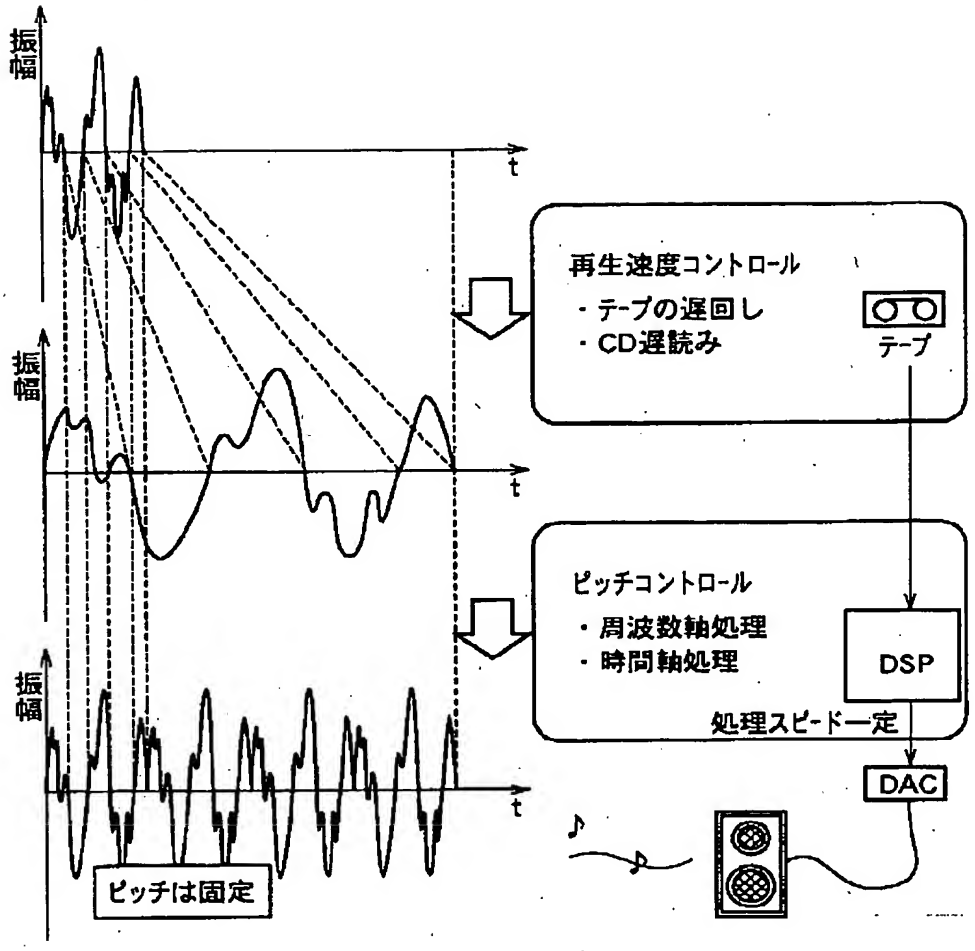
【図１３】


元のデータ
t
基本サイクル
ピッチダウンデータ
t
重複部分
ピッチアップデータ
t
欠落部分

【図１５】


power
t
ピッチは2倍
タイムは1/2倍
逆FFT
上位にシフト
X2
FFT
X1/2
ピッチは1/2倍
タイムは2倍
下位にシフト

【図１６】



フロントページの続き

(72)発明者　若杉　純
神奈川県川崎市幸区小向東芝町１番地　株式会社東芝マイクロエレクトロニクスセンター内

Ｆターム(参考)　5D045 BA01 BB02
5D108 BF06
5J064 AA01 BA09 BA16 BB04 BC07 BC11 BC16 BD02